PROCEEDINGS

55. IWK

Internationales Wissenschaftliches Kolloquium
International Scientific Colloquium

13 - 17 September 2010

Crossing Borders within the ABC

Automation,

Biomedical Engineering and

Computer Science

Faculty of
Computer Science and Automation

www.tu-ilmenau.de

TECHNISCHE UNIVERSITÄT
ILMENAU

Impressum
Published by

Publisher:	Rector of the Ilmenau University of Technology Univ.-Prof. Dr. rer. nat. habil. Dr. h. c. Prof. h. c. Peter Scharff
Editor:	Marketing Department (Phone: +49 3677 69-2520) Andrea Schneider (conferences@tu-ilmenau.de) Faculty of Computer Science and Automation (Phone: +49 3677 69-2860) Univ.-Prof. Dr.-Ing. habil. Jens Haueisen
Editorial Deadline:	20. August 2010
Implementation:	Ilmenau University of Technology Felix Böckelmann Philipp Schmidt

USB-Flash-Version.

Publishing House:	Verlag ISLE, Betriebsstätte des ISLE e.V. Werner-von-Siemens-Str. 16 98693 Ilmenau
Production:	CDA Datenträger Albrechts GmbH, 98529 Suhl/Albrechts
Order trough:	Marketing Department (+49 3677 69-2520) Andrea Schneider (conferences@tu-ilmenau.de)

ISBN: 978-3-938843-53-6 (USB-Flash Version)

Online-Version:

Publisher:	Universitätsbibliothek Ilmenau ilmedia Postfach 10 05 65 98684 Ilmenau

Home / Index:
http://www.db-thueringen.de/servlets/DocumentServlet?id=16739

RESOLUTION ENHANCEMENT AND NOISE REDUCTION OF INFRARED VIDEO STREAMS

Carsten Lucht, Karl-Heinz Franke

Ilmenau University of Technology
Computer Graphics Group
P.O. Box 100 565,
D-98684 Ilmenau, Germany

Rainer Jahn, Rico Nestler

Zentrum für Bild- und
Signalverarbeitung e. V.
Werner-von-Siemens-Str. 10,
D-98693 Ilmenau, Germany

ABSTRACT

The images of thermographic (IR) cameras suffer from relatively small resolution and strong stochastic noise. Hence, it is recommended to improve the image quality by subsequent processing steps. The so-called super-resolution is a well suited approach for image improvement of video streams. It exploits the fact, that adjacent images of a video stream contain the same object scene but they differ in "microscopic" details. These small differences can be used to multiply the image resolution. A super-resolution technique generates a new image from every few (typically four to ten) adjacent video images and consists of roughly three steps: image registration, alignment, and fusion. We have reworked, implemented, and tested some variants of super-resolution.

Index Terms— resolution enhancement, noise reduction, thermography, video stream

1. INTRODUCTION

The term "super-resolution image reconstruction" (or simply "super-resolution") denotes methods for increasing the image resolution beyond the limit that is given by the pixel grid. A single image does usually not contain enough information to increase the resolution in this kind. Super-resolution needs a sequence of images that slightly vary from frame to frame. The basic principle is known at least since 1984, as cited in [1]. It is mainly used in thermography because of the relatively small resolution of IR sensors.

In order to understand the super-resolution, we suppose that the object scene is fixed and the camera pans over the scene. So we obtain images with nearly the same content that are only translated and possibly rotated to each other. As far as the image translation is a multiple of the pixel distance and rotation is negligible, the objects are identical in all details at adjacent images. However, the translation amount actually involves a random fractional part. Hence the objects are sampled at distinct points in different images. If we combine these sample grids, we get a refined grid in the overlapping area of the scene. This happens even when only few images are used, see fig. 1.

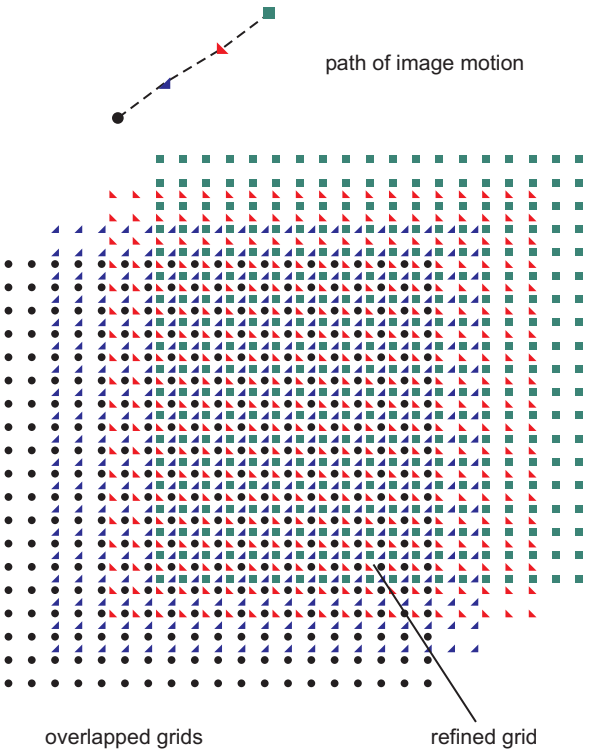

Fig. 1. Principle of super-resolution.

Instead of using random camera steps, one can generate regular image sensor shifts inside the camera. This approach, called "micro-scanning", is only suitable for cameras that are fixed on a tripod. The latter will not be discussed here.

In the next sections we will first explain the basic steps of super-resolution. Then we will treat one of these steps, the matching of adjacent images (registration). After this, the intrinsic image reconstruction will be investigated by means of two approaches of different complexity.

2. STEPS OF SUPER-RESOLUTION

The process of super-resolution is roughly illustrated in fig. 2. A real technique may deviate from this scheme in some details. This chart nearly represents the inversion of the real image formation process with one essential difference: We think the real object to be spatially band-limited, so that it can be replaced by a discret structure according to the sampling theorem. This structure is regarded as the desired high-resolution image.

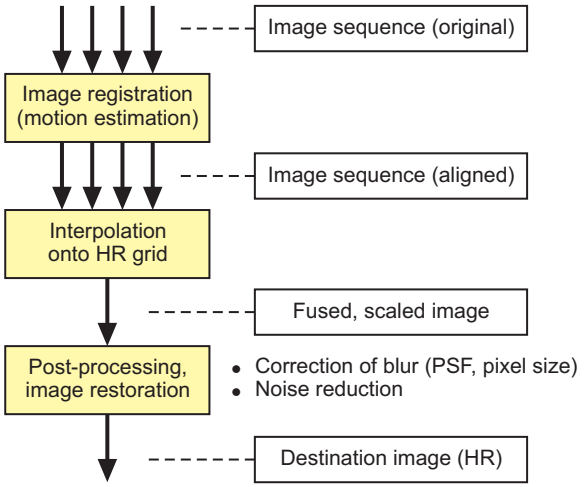

Fig. 2. Process of super-resolution, adapted from [2].

The first step, called "registration", means a technique for finding optimal coordinate transformations between images to align them. It is applied to every few adjacent images of the incoming stream. The output of the registration can be regarded as an aligned set of images that are as similar as possible. In the next step, this alignment is executed and the images are interpolated onto the given high-resolution grid.

Along with the interpolation, the images must be suitably combined to obtain an improved representation of the object scene. Thereby some serious problems occur. The source images may be blurred by optical aberrations ("point source transmittance" - PSF), so that higher spatial frequencies are lost. However, since the destination image is able to depict higher frequencies, it should contain corresponding information. To solve the problem, the higher spectral parts must be reconstructed from the source images.

Such reconstruction techniques commonly realise the inversion of an image formation step. But the equation system that has to be inverted is generally badconditioned. So the noise will prevent the inversion from giving a useful result. Therefore, so-called "regularisation" methods exist that improve the conditioning by additional constraints, like a smooth image intensity. In section 4.2 we will deal with that. Image interpolation and reconstruction or restoration cannot be strongly separated. The division into two steps according to fig. 2 is quite formal. In some variants, these steps may be partially exchanged.

3. IMAGE REGISTRATION

The aim of registration is to estimate the motion of objects between images. We assume that some adjacent frames of a video stream contain the same structures and differ only in the position of these structurs. Further we assume, there is a global geometric transformation that quantifies this motion by only few parameters. If the real objects are moving slowly with respect to the frame rate and the camera pans over the scene, this assumption holds.

An established method for determining these geometric parameters is a version of the Gauss-Newton iteration. The following brief overview is similar to that by Katsaggelos et al. [3]. We consider a short sequence of about four to ten images. One of these images is defined as reference image with a gray value function of $g_0(x, y)$. An arbitrary other one is regarded as "current image" with a gray value function of $g(x, y)$. We desire the geometric transform (X', Y') between the two images. (X', Y') provides the point (x', y') of the reference image that corresponds to a given point (x, y) of the current image:

$$x' = X'(x, y, \underline{w}), \quad y' = Y'(x, y, \underline{w}), \quad \forall\,(x, y). \tag{1}$$

There is an additional variable, the vector $\underline{w}$, that comprises the unknown geometric parameters.

If the transformation is correct, the following formula holds:

$$g(x, y) = g_0(X'(x, y, \underline{w}), Y'(x, y, \underline{w})), \quad \forall\,(x, y). \tag{2}$$

This can be regarded as equation system in $\underline{w}$. Since there are many more points (x, y) (one for each pixel) than elements in $\underline{w}$, the equation system is very overdeterminated. It cannot be solved exactly, but only in the sense of minimising a residual RMS error.

In order to handle the system (2), the Gauss-Newton iteration uses a linearisation $\tilde{g}$ of g, that is given by

$$\tilde{g}(x, y, \underline{w}) = g_0(x, y) + \sum_{k=1}^{q} h_k(x, y) \cdot w_k \tag{3}$$

with the abbreviation

$$\begin{aligned} h_k(x, y) = {} & \left.\frac{\partial g_0}{\partial x}\right|_{(x,y)} \cdot \left.\frac{\partial X'}{\partial w_k}\right|_{(x,y,\underline{0})} \\ & + \left.\frac{\partial g_0}{\partial y}\right|_{(x,y)} \cdot \left.\frac{\partial Y'}{\partial w_k}\right|_{(x,y,\underline{0})}. \end{aligned} \tag{4}$$

All terms in (4) can approximately be calculated. We obtain the derivations $\partial g_0/\partial x$ and $\partial g_0/\partial y$ by means

of appropriate discrete convolution operators, e.g. the Sobel operator. The other two derivations result from the ab initio known equations of X' and Y' as functions of the geometric parameters $\underline{w} = (w_1, w_2, \ldots)^T$. An estimation $\underline{\hat{w}}$ of $\underline{w}$ follows from the requirement to minimise the RMS error:

$$\underline{\hat{w}} = \arg\min_{\underline{w}} \left(\sum_{(x,y)\in\mathcal{I}} \left(\tilde{g}(x, y, \underline{w}) - g(x, y)\right)^2 \right). \tag{5}$$

The index set $\mathcal{I}$ comprises the positions of all or the most pixels of the images.

Eq. (5) forms a linear equation system in $\underline{w}$ that can be solved by inversion of the coefficient matrix. This matrix is regular in practice. Since the solution holds only for a linear approximation of the original task, it must be improved by an iterative process. For this, the current image is subjected to the inverse of the geometric transfom (X', Y'), using the calculated parameter vector. This yields a corrected image that should be more similar to the reference image. Now, the reference image is replaced by this corrected image, and the calculation of the parameter vector according to (5) is repeated. We get a new transform $(X', Y')_{new}$. Then, the (original) current image is subjected to the combined inverse transform ${(X', Y')_{new}}^{-1} \circ (X', Y')^{-1}$. This results in an corrected image, that is even more similar to the reference image. The latter one is replaced by the new corrected image, and so on. If the changes of the calculated parameter vector will fall below a given limit, the process will be finished.

According to many references, e.g. [1] and [4], we restrict the vector $\underline{w}$ to two translation components and one rotation component:

$$\underline{w} = (\Delta x, \Delta y, \varphi)^T. \tag{6}$$

The geometric transformation will then be expressed as follows:

$$\begin{aligned} X'(x, y, \underline{w}) &= x \cdot \cos\varphi - y \cdot \sin\varphi + \Delta x, \\ Y'(x, y, \underline{w}) &= x \cdot \sin\varphi + y \cdot \cos\varphi + \Delta y. \end{aligned} \tag{7}$$

Inversion and combination of this special transform can be simply calculated: the first means a swap of the parameter signs, and the latter means addition of the respective parameter vectors.

4. IMAGE RECONSTRUCTION

4.1. Pure interpolation

The simplest way to realise the middle and last step in fig. 2 is to interpolate and superpose the camera images onto the high-resolution grid. Interpolation means that the pixel positions are scaled and remapped at the HR grid according to the registration result. Superposition means, the interpolated images are pixelwise averaged. This fast method is suitable for real-time applications, however, the potential to improve the images is strongly limited. It is not possible to effectively reduce blur and artifacts from the final HR image. So the actual aim of increasing the resolution may not be fully satisfied. But there is a second effect of all super-resolution techniques: The small-scaled noise patterns are noticeably reduced.

We have tested this method with images provided by an IR camera. The original size of 320 x 240 pixels was doubled upto the HR size of 640 x 480 pixels. Fig. 3 shows the first, fourth and tenth image of a sequence taken from a video stream. The camera was hand-held and slowly moved over an almost still scene. In the interpolation step, we have used the bilinear approach. Trials with bicubic interpolation have not significantly improved the result.

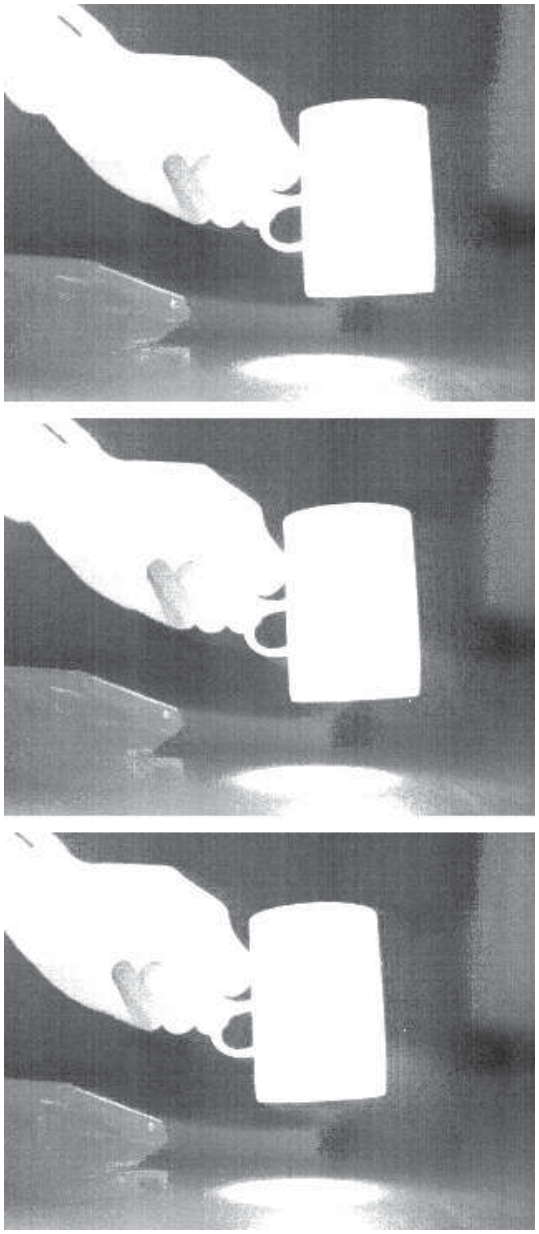

Fig. 3. Test images No. 1, 4, and 10 of an IR video stream.

The more camera images are processed to one HR image, the more details and fewer noise should be contained in the HR image. This holds if the objects stay inside the field of view and the registration works exactly. In our experience, the optimum is reached if four till ten images are fused. An example is shown in fig. 4. Above the HR image from the first four test images is shown. Below we see the result from all ten test images.

Fig. 4. Super-resolution result from four images (above) and ten images (below) of the test sequence, respectively. The "pure interpolation" method was applied.

The square label in both images encloses an area that is detailed in fig. 5. For comparision, the same area of the first unprocessed test image is shown above in fig. 5. This image was only extended to the HR grid by bilinear interpolation. In the other views, the results of super-resolution are shown. The middle view relates to fusion of four images and the lower view to fusion of ten images. We find that the image quality (resolution and noise suppression) is increased from up to down.

In the lower left corner of the resulting images a remarkable motion blur appears. This is caused by a slow shift of the cup relative to the other objects. Nevertheless, the registration works well and focuses on the greatest inflexible object contained in the field of view. If this solution is not sufficient, the local optical flow must be calculated in a different way. A suitable method is the so-called hierarchical block matching. This means a block matching at which the correlation blocks and step sizes are successively refined upto a subpixel level. We have tried this approach, but it operates too slowly for a real-time application.

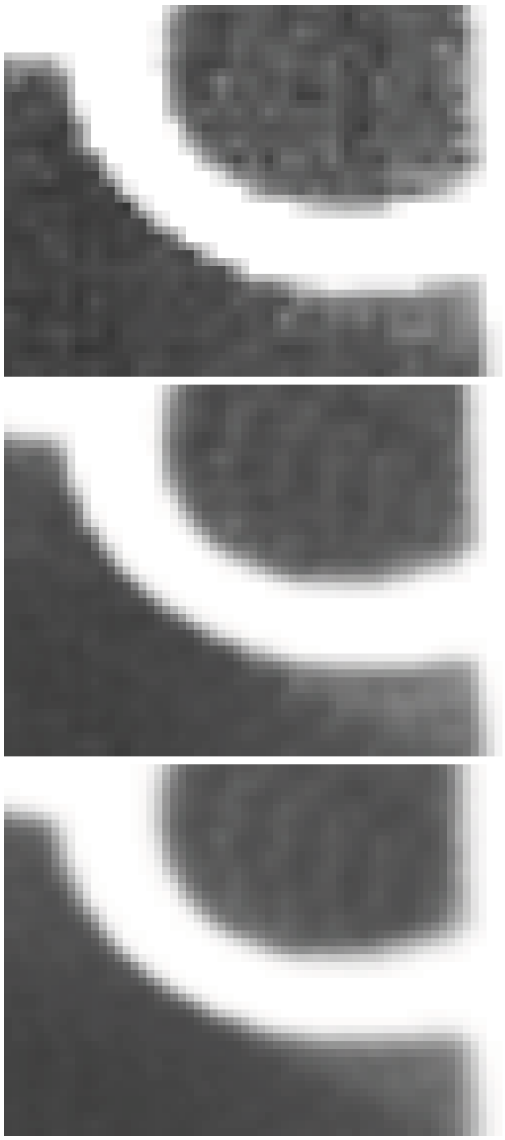

Fig. 5. Zoomed views (mid and below) of the marked areas in fig. 4. Above, the same area of the first test image is shown, that is only extended to the HR grid.

4.2. Regularised image restoration

The above presented method of super-resolution has a serious disadvantage. It is assumed that the object scene is perfectly projected on the image sensor, and physical aspects of the image formation are not regarded. In this section we briefly describe an improved approach that encloses some of these aspects.

For most purposes, one can assume that a general linear relation exists between the ideal HR image and

each camera image. It was already explained that the HR image can be identified with the real object scene in some sense, see section 2. Suppose, we combine p camera images to one HR image with the gay value function $\underline{f}$. The gray value functions of the camera images may be $\underline{g}_1, \ldots \underline{g}_p$. Then the following observation model is valid:

$$\underline{g}_k = \underline{\underline{M}}_k \cdot \underline{f} + \underline{n}_k, \qquad k \in \{1, 2, \ldots, p\}. \quad (8)$$

In order to get the most general linear form, each of the images $\underline{g}_k$ and $\underline{f}$ must be expressed as a single column vector, and the transformation matrix $\underline{\underline{M}}_k$ must have the appropriate size. The vector $\underline{n}_k$ represents the pixelwise additive noise contribution of the image $\underline{g}_k$. Due to their large size, the terms in (8) cannot be handled in this general kind. This equation and the following ones are only formally valid. The terms must be simplified to the most essential aspects of the observation model.

The transformation matrix can be expressed as product of matrices:

$$\underline{\underline{M}}_k = \underline{\underline{D}} \cdot \underline{\underline{B}}_k \cdot \underline{\underline{T}}_k \quad (9)$$

Here, $\underline{\underline{D}}$ describes the sampling process of the image sensor, $\underline{\underline{B}}_k$ the blur by optics and pixel aperture, and $\underline{\underline{T}}_k$ the motion and geometric transform according to the registration result.

In principle, the HR image can be obtained by solving a linear equation system. If a proper number of p equations (8) are combined to a single system and zero noise vectors are assumed, we get a well-defined solution for $\underline{f}$. In our case, it is $p = 4$ and the vector $\underline{f}$ is four times as long as a single vector $\underline{g}_k$. Unfortunately, this way is mostly not useful since noise and computational artifacts would be fully transmitted into the solution.

Therefore, a regularisation term should be introduced. Hardie [4], Park [2] and others propose to calculate the desired HR image $\underline{\hat{f}}$ as follows:

$$\underline{\hat{f}} = \arg\min_{\underline{f}} \left(\sum_{k=1}^{p} \left\| \underline{\underline{M}}_k \cdot \underline{f} - \underline{g}_k \right\|^2 + \lambda \cdot \left\| \underline{\underline{C}} \cdot \underline{f} \right\|^2 \right). \quad (10)$$

The regularisation matrix $\underline{\underline{C}}$ defines the constraints and may be a kind of high-pass filter. A simple realisation is given by

$$C_{mn} = \begin{cases} 1 & n = m, \\ -1/4 & n \text{ neighbour point of } m, \\ 0 & \text{else.} \end{cases} \quad (11)$$

The parameter λ is called "damping term". It controls the influence of the constraints on the resulting image. If the smallest value of $\lambda = 0$ is used, the regularisation ist suppressed and we get the mentioned direct solution with its insufficiencies. In the opposite case, if λ would grow to infinity, the resulting image would totally lose its specific content.

Eq. (10) describes a multilinear regression task that can be expressed in closed form as linear equation system. Due to its great size, it is impractical to solve this system by a standard method. Instead, an iterative splitting approach is recommended.

Our experimental results with this approach are described now. We have implemented the so-called Richardson algorithm as explained in [5] and other sources to solve the equation system (10) iteratively. The appropriate damping value λ cannot be calculated by a simple formula, so we have tried some values between 0 and 10. An example of the results is shown in fig. 6. In all subviews the calculated HR image from the first four camera images is illustrated. For better visibility, the views are restricted to the same image detail as in fig. 5. From up to down the results for $\lambda =$ 0, 1, 10 are presented, respectively.

We have simplified the transformation matrices $\underline{\underline{M}}_k$ as follows: The motion was expressed as global translation and rotation, and the blurring effect was regarded as convolution with the optical PSF and the square pixel aperture. The PSF was modeled as a Gaussian distribution, and its width was estimated from edges that are contained in the camera images.

In the upper view of fig. 6 the case of suppressed regularisation (damping) is shown. As expected, the image is disturbed by artifacts and noise, but the sharpness is higher than in the previous reconstruction method. The mid view shows the case of moderate damping. One can see that the high-frequency disturbance is mitigated. Unfortunately, this leads to somewhat lower edge sharpness. The result is now comparable to that of the previous method, see the mid view of fig. 5. In the lower view we see the effect of a stronger damping. The disturbance is further reduced, but the lost of sharpness is no longer acceptable. Our experiments have shown that the optimal value of λ lies in the interval $[0.5, 2.0]$.

5. CONCLUSIONS

In this paper, the principle of super-resolution was described, and related own investigations were presented. It could be shown that generating of super-resolved images from IR video sequences works quite well. First, a simpler technique was presented that is suitable for real-time applications. This method provides images, that are visually sharper and cleaner than the originals. But the effect of resolution enhancement is not so obvious. Nevertheless, we have implemented this method in a stand-alone software that can be considered as assessory of an IR camera. It processes the source images offline, but the operation speed on an up-to-date PC allows a nearly real-time processing.

Second, a more sophisticated approach of super-resolution was investigated. This technique considers some physical aspects of image formation. The pro-

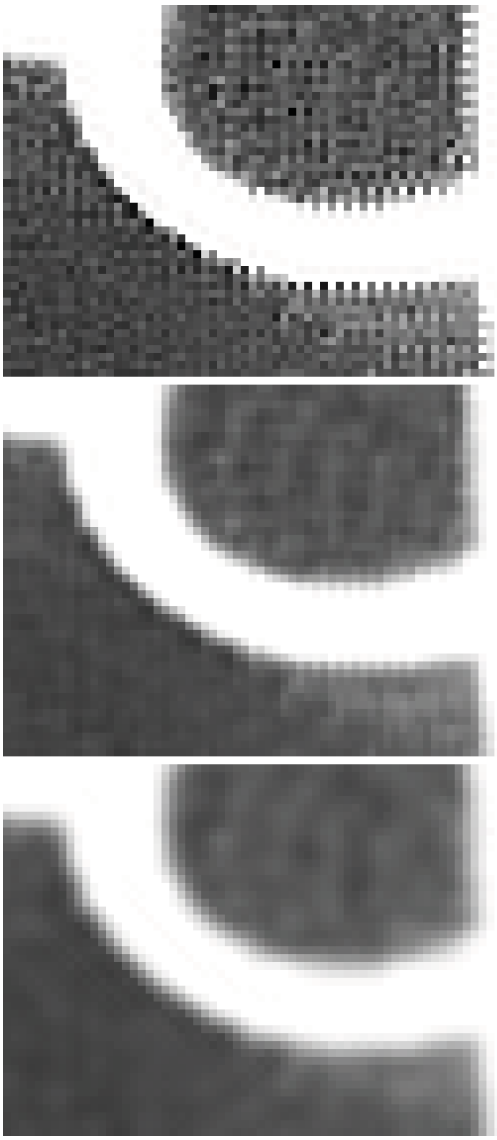

Fig. 6. Super-resolution result from four images of the test sequence. The ”regularised restoration” method was applied.

cess can be parametrised to get either quite sharp and somewhat noisy or less sharp and cleaner HR images. Since the kernel process is iterative, the operation time is much longer than that of the first method. Real-time ability cannot be achieved by PC technology. But it would be promising to implement the kernel algorithm on an embedded system or on a Graphics Processing Unit.

6. ACKNOWLEDGMENTS

The content of this paper is part of the research project NEMO, supported by the Free State of Thuringia, Germany. One of the project partners, the ESW GmbH, has provided source images from a VarioCAM® IR camera.

7. REFERENCES

[1] D. Keren, S. Peleg, and R. Brada, “Image sequence enhancement using sub-pixel displacements,” in *IEEE Conference on Computer Vision and Pattern Recognition*, Ann Arbor, Juni 1988, pp. 742–746.

[2] S. C. Park, M. K. Park, and M. G. Kang, “Super-resolution image reconstruction: a technical overview,” *IEEE Signal Processing Magazine*, pp. 21–36, Mai 2003.

[3] A. Katsaggelos, R. Molina, and J. Mateos, *Super Resolution of Images and Video*, Morgan and Claypool Publishers, 2007.

[4] R. Hardie et al., “High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrafred imaging system,” *Optical Engineering*, vol. 37, no. 1, pp. 247–260, 1998.

[5] J. J. Dongarra, *Numerical linear algebra for high performance computers*, Society for Industrial and Applied Mathematics, 1998.